DOCUMENT RESUME

ED 178 546 TM 007 132

AUTHOR	Hartsough, Carolyn S.; And Others
TITLE	The Validation of a Nonintellectual Assessment Procedure for the Early Screening of Gifted and EMR Pupils.
INSTITUTION	Educational Testing Service, Princeton, N.J.
SPONS AGENCY	California State Dept. of Education, Sacramento.
REPORT NO	ETS-PR-77-15
PUB DATE	[77]
GRANT	76-64-G
NOTE	57p.
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	*Academically Gifted; Academically Handicapped; *Behavior Rating Scales; Culture Free Tests; *Educable Mentally Handicapped; *Informal Assessment; Intelligence Quotient; Peer Evaluation; Primary Education; *Screening Tests; Self Evaluation; Student Evaluation; Test Bias; *Test Validity

ABSTRACT

The validity of a screening process for the early identification of gifted and educable mentally retarded children was examined. The process was originally developed by Lambert and Bower, for the identification of learning handicaps. Twenty-five kindergarten and first grade classrooms were studied. The information collected included the teacher, peer, and self ratings in the screening process, a measure of aptitude, and student characteristics--sex, socioeconomic status, ethnic group, and bilingual status. Objectives were to: study the feasibility of using the assessment materials for initial screening of gifted and of educable mentally retarded pupils; extend teachers' understanding of the emotional and academic needs of their students; and compare the effectiveness of the assessment process with existing information which might be used for screening purposes. Results indicated that: (1) the screening materials tended to "overpredict" or include too many children as candidates for program placement when compared to IQ scores; (2) teachers tended to "underpredict" or exclude too many children as candidates for program placement when compared to IQ scores; and (3) the screening materials appeared to be less vulnerable to socioeconomic status and ethnic or bilingual background than IQ scores. (Author/GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED178546

PR-77-15

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

The Validation of a Nonintellectual Assessment Procedure for the Early Screening of Gifted and EMR Pupils

Carolyn S. Hartsough

University of California
Berkeley

Patricia Elias
And
Patricia Wheeler

Educational Testing Service
Berkeley

TM007 132

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
P. Wheeler
ETS
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

EDUCATIONAL TESTING SERVICE
PRINCETON, NEW JERSEY

FOREWORD

The study reported here was part of a program of research in Special Education in the Research and Advisory Services Department of Educational Testing Service.

The study was carried out during the academic year 1976-77 under the auspices of the Special Education Research Program, supported by Grant No. 76-6[illegible]-G between the State Department of Education and Educational Testing Service.

The full report is reproduced here in this form for distribution as a technical report under the grant, and in order to make complete findings available for others engaged in this research area. Results of this study are the sole responsibility of the investigators. Official endorsement of the California State Department of Education or Educational Testing Service is not implied.

Patricia Elias
Patricia Elias
Educational Testing Service, Berkeley

Patricia Wheeler
Patricia Wheeler
Educational Testing Service, Berkeley

Carolyn S. Hartsough
Carolyn S. Hartsough
University of California, Berkeley

ACKNOWLEDGEMENTS

Special thanks to: participating teachers and their principals; district and county personnel: Paul M. Hillar, Walter V. Klas, John Sugiyama, Anne Bradfield, Fred Kjelland, Elizabeth Elcock; to the developer of the materials for support, suggestions and leadership in inservice meetings, Nadine Lambert, U.C.B.; examiners: Mary Corder, Patricia Taylor, Michael Walsh, Ruth Antonini, Barbara Walker, and Kathleen Geiger; translators: Carmen Windhorst (Spanish) and Teresita Andal Young (Tagalog); and to Alice Setteducati and DeeAnn Cuthbertson for editing and typing; and to John Stevenson and Becky Yheulon for printing and binding.

Carolyn S. Hartsough
Patricia Elias
Patricia Wheeler

ABSTRACT

Educational Testing Service's Berkeley (CA) office was funded by the California State Department of Education in 1976-77 to test the validity of a screening process, "The Nonintellectual Assessment of Effective School Behavior," for the early identification of gifted and educable mentally retarded children. The process was developed by Nadine M. Lambert and Eli M. Bower of the University of California at Berkeley for use in identifying learning handicapped children.

The target population was the teachers in 25 kindergarten and first grade classrooms in five Northern California school districts (Alameda City, Ceres, Empire, Modesto, and Mt. Diablo) and their pupils. The information collected included the teacher, peer, and self ratings in the screening process, a measure of aptitude, and demographic data about the pupils.

The objectives of the study were:

- To study the feasibility of using the assessment materials for initial screening of gifted and of educable mentally retarded pupils.
- To extend teachers' understanding of the emotional and academic needs of each child in their classrooms.
- To compare the effectiveness of the assessment process with existing information which might be used for screening purposes.

The results were:

- The screening materials tend to "overpredict" or include too many children as candidates for gifted and educable mentally retarded program placement when compared to IQ scores.
- Teachers tend to "underpredict" or exclude too many children as candidates for gifted and educable mentally retarded program placement when compared to IQ scores.
- The screening materials appear to be less vulnerable to socioeconomic status and ethnic or bilingual background than IQ scores.

TABLE OF CONTENTS

	Page
Foreword	i
Acknowledgements	ii
Abstract	iii
List of Tables	v
Introduction	1
Purposes	2
Historical Overview	4
Concurrent Validity	4
Long-term Predictive Validity	5
Methods	7
Sample	7
Instrumentation	8
Procedures	13
Teacher Feedback	15
Data Analysis and Results	17
Data Analysis	18
Description of multiple cut-off procedure for screening	19
Analytic strategy for validity study for gifted and educable mentally retarded	20
Results of the Validity Substudy	21
Forming IQ Criterion Groups	21
Discriminant analysis results	24
Results of the Screening Substudy	29
Incidental Findings	36
Congruence of teacher perception and project results	36
Comparison of results of screening with teacher nominations made prior to administration of materials	37
Comparison of results of administration of ability test in native language and in English	38
Summary and Recommendations	41
Conclusions	45
References	47

LIST OF TABLES

Table		Page
1	SES Level of Participating Children	9
2	Ethnic Group of Participating Children	9
3	Dominant Language of Participating Children	10
4	Descriptive Statistics for Slosson Intelligence Test . .	22
5	Distribution of Pupils Obtaining Low and High IQ Scores Summarized by Sex, Socioeconomic Level, Ethnicity and Bilingual Status	25
6	Discriminant Function Analysis of "Nonintellectual Assessment" Measures for Low IQ Groups vs. Others . . .	26
7	Discriminant Function Analysis of "Nonintellectual Assessment" Measures for High IQ Groups vs. Others . .	27
8	Distribution of Pupils Screened by Teacher-Peer-Self Multiple Cut-off Criterion Categorized by Sex, Socioeconomic Level, Ethnicity and Bilingual Status . .	30
9	Distribution of Pupils Screened by Teacher-Peer Multiple Cut-off Criterion Categorized by Sex, Socioeconomic Level, Ethnicity and Bilingual Status . .	31
10	Distribution of Pupils Screened by Teacher-Self Multiple Cut-off Criterion Categorized by Sex, Socioeconomic Level, Ethnicity and Bilingual Status . .	32
11	Distribution of Pupils Screened by Peer-Self Multiple Cut-off Criterion Categorized by Sex, Socioeconomic Level, Ethnicity and Bilingual Status . .	33
12	Frequencies of Pupils Screened by Any of Multiple Cut-off Criteria Summarized by Major Demographic Variables	35
13	Scores Obtained for Pupils Administered Ability Test in English and in Tagalog	39

Introduction

Most school districts in California presently rely on teacher referral for selecting candidates in the various categories of exceptional children. Historically, this procedure has resulted in delays in the identification of such children until their fourth or fifth year in school. There are several reasons for this, including teacher attitudes and time constraints on the special services staff. By the time pupils are identified, however, these children have often experienced several years of failure or frustration in school, and have often developed fairly severe behavioral and emotional problems. Remediation then involves treating not only the special learning needs of these children, but also problems of motivation, self-concept, and relationships with peers. The instructional needs of children, as well as school goals, would be better served by the early identification of children who are in need of special learning assistance.

In recognition of the need for improved identification procedures, the California State Department of Education funded a study by Educational Testing Service which was designed to validate the efficacy of nonintellectual assessment materials for the early screening of gifted and educable mentally retarded (EMR) pupils. The screening materials examined in this study have demonstrated validity for the early identification of learning handicapped children. Incidental findings in related studies indicated that the procedure might have some potential in the screening of both gifted and EMR pupils.

The proposed procedure* is called "The Nonintellectual Assessment of Effective School Behavior" (Bower and Lambert, 1961). The instruments employed in the screening process were designed, developed, and used for the purpose of identifying those children whose emotional and behavioral symptoms are associated with learning handicaps. Three instruments are used to measure how each child in a regular classroom is perceived by the teacher, by peers, and by himself/herself.

Purposes

Both the original research and later longitudinal studies with the nonintellectual assessment procedures document its concurrent and long-term predictive validity as a more efficient strategy than the referral process for the identification of children with a variety of learning related handicaps. The purposes of this study were:

1. The evaluation of the nonintellectual assessment materials as procedures for the preliminary "screening" or initial identification of gifted and educable mentally retarded pupils, with particular emphasis on culturally different pupils.
2. The comparison of the efficiency of the assessment procedures with teacher judgment at the end of the school year.
3. The extension of teachers' understanding of the emotional and academic needs of each pupil in their classrooms.

In addition, we will use the full pupil population to describe the learning handicapped children identified by the screening process.

*Formerly "A Process for In-School Screening of Emotionally Handicapped Children".

Three substudies will be discussed: congruence of teacher perception and project results, comparison of results of screening with teacher nominations made prior to administration of materials, and comparison of results of administration of ability test in native language and in English.

Historical Overview

Concurrent Validity

Although the problem of establishing suitable criteria for concurrent validation of the screening process, as for any new measurement system, was a major obstacle, the technical report (Bower and Lambert, 1961b) for the materials presented six such investigations. A later study by Lambert (1963) showed that the screening procedure is valid for predicting independently derived clinical assessments of school adjustment. As a totality they support the hypothesis that the screening process identifies children with a variety of learning-related behavioral handicaps more economically and efficiently than the teacher referral process.

The following study (Bower and Lambert, 1961b) is illustrative. In the spring of 1958, selected school districts were asked to list names of pupils thought by the administrative office and guidance personnel to be behaviorally handicapped. Names were listed and later checked with lists of children identified in the screening process by two out of three ratings at the negative extremes. Teacher, peer and self ratings on a school population of 2,198 children in grades 3, 4, and 5 in three school districts were obtained. Guidance personnel nominated 150 pupils (7 percent) as behaviorally handicapped. Of the 140 who were present when the screening materials were administered, 96 percent of these children (or 135 of 140) were screened by the recommended multiple cut-off procedure. Furthermore, using the same criteria of two out of three ratings at the negative extremes of the distributions, an additional 7 percent (or 150 of the 2,198 children) from the total population was screened. This suggested that at least one-half of the children who are at risk of educational failure are probably not noted or observed by the usual referral process.

Long-term Predictive Validity

Continued research (Lambert, 1972) has shown that these screening materials are effective in predicting long-term outcomes as well as concurrent status. Children who were screened in second grade and fifth grade were assessed again in twelfth grade. Results showed that the combined ratings were significant predictors of adjustment (as judged by high school counselors), GPA, delinquency and graduation. Furthermore, the magnitude of these correlations exceeded the more traditional indicators of successful status, such as IQ and achievement scores. Case histories of twelfth graders who would have been identified by the multiple criteria of the screening procedure in elementary school revealed a multitude of social, emotional and academic problems. For example, nine twelfth-grade boys out of the total of 56 twelfth graders who met the screening criteria seven years earlier could be located at a follow-up. By the end of high school, four of the nine had delinquency records, three others were referred to the guidance office, one of these three dropped out of school, and another was referred for remedial help. With the exception of one pupil, all of the others had a mean grade point average of 2.1 or lower on a four-point scale for twelfth grade course work. Moreover, all these pupils who had been identified earlier were still manifesting school and personal difficulties.

Work was recently completed which served, in effect, as a pilot test for the present project. Approximately 10 teachers and 300 children participated. The exploration of the use of the screening materials for gifted and EMR screening did not yield definitive patterns. Certain

trends did appear, however, which are similar to those that emerged in this study (Lambert, Hartsough, and Urbanski, 1977). Some specific comparisons will be made later in the report.

Methods

Sample

For the purposes of a validation study, it was not necessary to obtain a random sample. However, the sample did need to represent a variety of types of districts and schools with ethnically and culturally diverse student populations.

A total of eight Northern California districts and one county (representing 23 districts) were invited to participate in the project. Two districts and the county (three districts) agreed to participate. We then contacted school personnel and invited the kindergarten and first grade teachers at each school site to participate. This process required considerably more time than anticipated. It took from mid-July through early February to identify our volunteer sample of 25 teachers in 14 schools. Thirteen taught kindergarten; eight, first grade; and four, combination classes (kindergarten/first and first/second).

The districts which participated included Alameda City, Ceres, Empire, Modesto City, and Mount Diablo. They ranged in size from 1,343 to 42,894 (California State Department of Education, 1977) and covered urban, suburban and rural communities. According to the data in the California State Testing Program, 1971-72 and 1972-73, Profiles of School District Performance, the minority pupil enrollment in these districts in 1971-72 ranged from 6.0 percent to 22.4 percent. The state percentile ranks of the district's poverty index in 1972-73 ranged from 23 percent to 92 percent. And the state percentile ranks of the district's grade 6 scholastic aptitude level ranged from 12 percent to 79 percent.

A total of 536 pupils in the 25 classrooms participated in the project. Parental permission was not obtained for an additional 73 pupils in these classrooms. About 88 percent of the parents gave permission for their children to take part in this study. Forty-six percent of the sample were girls and 54 percent, boys. Tables 1, 2, and 3 contain data on the socioeconomic status level, the ethnic group, and the dominant language of the participating children. The sample included children from all SES levels, with most at the "lower middle" level. Over 27 percent of the children were from ethnic minority groups. English was the second language for 13 percent of the sample. The sample was ethnically and culturally diverse.

Instrumentation

"Nonintellectual Assessment of Effective School Behavior" consists of three instruments: the Pupil Behavior Rating Scale, the Who Could This Be Game, and A Picture Game. These instruments were completed for each participating child as described in the Procedures section. In addition, each child was administered the Slosson Intelligence Test (SIT).

The screening process was developed by Eli Bower and Nadine Lambert of the University of California at Berkeley (Bower and Lambert, 1961a, 1961b) for use in identifying children whose emotional and behavioral symptoms are associated with learning handicaps. The version used in the present study is a revision of the earlier process, done by Dr. Lambert (Lambert, 1974). It is distributed by the School of Education at the University of California, Berkeley.

Table 1

SES Level of Participating Children

Category	Number	Percent
Upper	47	8.8
Upper Middle	93	17.4
Middle	83	15.5
Lower Middle	217	40.5
Lower	70	13.1
Not reported	26	4.9

Table 2

Ethnic Group of Participating Children

Group	Number	Percent
White	389	72.6
Black	19	3.5
Chinese	11	2.1
Filipino	47	8.8
Japanese	7	1.3
Mexican	43	8.0
Other, Not reported	20	3.8

Table 3

Dominant Language of Participating Children

Language	Number	Percent
English	466	86.9
Chinese	4	0.7
Japanese	2	0.4
Native American	1	0.2
Spanish	29	5.4
Tagalog, Filipino	24	4.5
Other, Not reported	10	1.9

"The Pupil Behavior Rating Scale," a teacher rating instrument, was designed to assess behavioral attributes associated with school adjustment. A child's rating on an attribute is based on the teacher's assessment of behavior and the teacher's judgment of the child's relative position on a well-delineated continuum marked by descriptive statements. The scale consists of eleven items thought to tap the important nonintellectual attributes which contribute to successful or unsuccessful functioning in the classroom. Each child is ranked on a scale from 0 to 3, divided into tenths, from "least like" to "most like" on each attribute. In order to help teachers anchor the ratings and to avoid teacher bias, each item has several descriptive behaviors along its continuum to cue the teacher to the degree of behavior associated with the value assigned to the pupil.

The "Who Could This Be Game" is a peer rating instrument for use in kindergarten through grade 3. It was developed as a means of analyzing, in a systematic and measurable way, how children are perceived by their peers. It is composed of eleven picture cards with a total of 20 items. Five pictures are of boys and five are of girls engaging in positive or neutral behavior. Five pictures are of boys and five are of girls engaging in problem or negative behavior. Each pupil is asked to indicate which classmates are associated with each kind of behavior. The total frequency with which each pupil is mentioned by other pupils (visibility), as well as the frequency with which each student is mentioned by peers in association with positive and negative behavior roles, are the primary data sources used in applying the results to the screening process. Children are queried individually and their responses recorded by the examiner.

"A Picture Game" was designed to give a measure of young children's perception of self. It consists of 70 pictures plus two sample pictures. Each picture is illustrative of normal home and school relationships and events. With the exception of the two sample pictures and the first ten pictures, each picture is emotionally neutral in the portrayal of the relationship or event. The child is asked to classify each picture into one of two categories, "This is a happy picture" or "This is a sad picture." The classification is done by marking a happy or sad face shown at the side of each picture. The child categorizes each picture in accordance with his/her perception of it. The first ten pictures the child judges are stereotypes, that is, obviously happy or obviously sad situations. The purpose of including them in the test items is to check on each child's understanding of the task. If the child judges the first ten pictures correctly, one can be fairly certain that he/she understands the process well enough to yield a valid score for screening. There are separate forms of the "Picture Game" for boys and girls.

In order to validate the process for screening of gifted and EMR children, it was necessary to administer an intelligence test to participating children. The Stanford-Binet and the Wechsler Intelligence Scale for Children--Revised (WISC-R) are the tests most commonly used by California districts for identifying gifted and EMR pupils. However, the relatively long period of time needed for administration limited their utility for our purposes in this project. In order to meet our time constraints and yet produce a measure of scholastic aptitude comparable to that obtained with the standard tests, an alternative measure needed

to be selected. Project staff reviewed several possible instruments for use in the study and selected the Slosson Intelligence Test, since it was felt it would provide a reliable and valid proxy for the instruments most commonly used to identify gifted and EMR pupils in California. This is an individually administered test which requires about 20-35 minutes to administer and is highly correlated with scores on the Stanford-Binet (S-B) and the Wechsler intelligence scales. For the S-B, the range is .76 to .90, and for the Wechsler, .54 to .93 (Buros, 1972).

Procedures

Following the identification of the sample, two activities were undertaken simultaneously: 1) the hiring and training of examiners, and 2) the inservice training of teachers. A total of nine examiners were selected for the project. Two were assigned to adapt the nonintellectural assessment procedures and the SIT for children whose dominant language was Spanish or Tagalog. All the examiners had had experience in testing and working with young children prior to the project. The Spanish and Tagalog examiners had often been called upon to do translations for other projects and were well qualified for the task at hand. All examiners were trained to administer the peer ratings and the SIT by Dr. Hartsough.

Two inservice training sessions were conducted, one in Stanislaus County in late January and one in Alameda City in early February. Not all participating teachers were able to attend one of these two sessions and some were provided inservice training individually by project staff. The main purposes of the inservice sessions were to explain the

assessment procedures and to acquaint the teachers with the issues under study. The sessions were conducted by Dr. Lambert and Dr. Hartsough. The topics included:

- an overview of the assessment process
- methods by which results could be summarized
- implications for classroom and outside intervention based on the findings of the screening process.

Following the inservice sessions, teachers had the option to withdraw from the project. All but a few teachers chose to participate.

The next major activity was to obtain parental permission. ETS developed forms for use in each district. Prior to distribution in the project classrooms, the forms were approved by district personnel. Both English and Spanish versions of a parent permission form were provided. The forms were sent home to parents with the children and returned to the teachers. As indicated earlier in this report, parental consent was granted for 536 of 609 children in project classrooms, an 88 percent agreement rate.

When all forms were returned by parents, project examiners worked with each teacher to schedule the testing of children in the classroom. The data collection activities occurred during March, April, and May. The participating teachers completed the teacher ratings, provided demographic data on each child, and administered the self ratings to groups of children in their classrooms. The project examiners administered the peer rating and the SIT on an individual basis to each child. Nine children were tested in Spanish, 13 in Tagalog and English, and the remainder in English.

Following the completion of data collection activities in each classroom, project personnel at ETS coded and scored each instrument. The data were then submitted to data processing personnel at UC Berkeley for keypunching and analysis. The data were analyzed and summarized within each classroom and across all 25 classrooms. The summarization by classroom allowed project personnel the opportunity to provide each participating teacher with feedback on their pupils prior to the end of the school year and to ascertain the congruence between our results and teacher perceptions.

Teacher Feedback

Each participating teacher was presented a summary of the results for her class on all the measures used in the project. The three screening instruments (teacher, peer, and self ratings) were combined via the multiple cut-off procedures described in the Data Analysis section. This procedure yielded a list of children who were suggested to the teacher as being vulnerable to potential behavioral and/or learning difficulties. In addition, we reported about children who were experiencing few difficulties in the view of the teacher and by the reports of peers. Also listed were those children who were "invisible" to their peers, that is, those who received no nominations or only one from peers, even for positive roles on the peer rating. In reporting results of the administration of the Slosson Intelligence Test, we noted those who achieved extremely high scores (145 and above), very good scores (130 to 144) or quite low scores (81 and below). For the reports

of IQ results, specific scores were not given. Results were presented in terms of ranges of scores and with descriptive terms, such as "excellent ability," "very good ability," or "ability indicating possible learning problems." Our purpose in such reporting was two-fold. We wished not only to abide by our agreement to inform participating teachers of results, but also to obtain a crude measure of consensual validity for the screening instruments by determining the extent of congruence between project results and teacher perceptions.

Data Analysis and Results

The primary purpose of this project was an evaluation of the "Nonintellectual Assessment" procedures for the preliminary screening of gifted and EMR pupils. One major set of results to be reported are related to this purpose. The procedures are, however, commonly used in a more general way to screen for a group of pupils whose emotional and behavioral symptoms predispose them to less extreme forms of learning handicaps (i.e., educationally handicapped). Our agreement with the five districts in which we worked specified that we would provide feedback to teachers about the results of the screening as it related to this more traditional use of the instruments. Since such information was already available as a by-product of this agreement, we felt it would enhance the value of the report as well as provide important information not generally obtainable to present descriptive statistics about the frequency and nature of the group of pupils screened in this way. Therefore, another major set of results from the investigation pertains to this outcome.

Another outcome from the teacher feedback phase of the study is of interest. In reporting back to teachers about the types of problems noted for the screened pupils, we requested that they indicate agreement or disagreement with our findings. One section of the report will describe the degree of congruence between our results and teacher perceptions.

Comparison of the screening process outcomes with teacher nominations of candidates for special education programs was another component of the study. Comparisons were made for three groups of candidates nominated by teachers prior to the testing: gifted, EMR, and learning handicapped.

Another by-product of the investigation not originally anticipated in the formulation of objectives, but of interest, is information which came from the administration of the criterion instrument, the Slosson Intelligence Test. A small subpopulation of children in both Stanislaus County and Alameda City were not proficient speakers of English. Accordingly, arrangements were made to administer the SIT and the other instruments in the native language. In Stanislaus County, a small sample of children were tested in Spanish. In Alameda City, the Tagalog speaking children were administered the test in both Tagalog and English. Having these data available, we will comment on the comparability of scores obtained with the administration of the SIT both in Tagalog and in English.

Data Analysis

Using the APPLE information storage and retrieval system (Lambert, et al., 1971), a computer file was built containing data gathered on those children with parent permission. The data were grouped by classroom. With the data stored in this fashion, various sorts of analyses were possible. In its traditional uses, scores from the three instruments of the "Nonintellectual Assessment" battery are combined via a multiple cut-off procedure, teacher by teacher, to indicate within each classroom those pupils who are at risk of learning handicaps. Such a procedure usually identifies or "screens" two to five children in each classroom who may need further evaluation by appropriate personnel, often the school psychologist. This cut-off procedure, to be described below, was used to obtain data to be reported in the section on the results of the screening substudy (p. 29).

For the research investigation of the validity of the procedures for screening of the gifted and of the educable mentally retarded (p. 20), a more powerful analytic strategy was needed and required that data be treated in a different way. Since the goal of this portion of the study was to determine whether any one or some subset of the scores from among those available from the screening instruments might have predictive value, use of the conventional multiple cut-off strategy would mean loss of information about the potential utility of the individual measures. To provide information about the individual instruments, an analytic strategy was chosen which required that the distribution of scores for the entire sample of pupils be analyzed. Accordingly, for the results reported in the validity section the data were aggregated across teachers.

Description of multiple cut-off procedure for screening. In order to screen children for learning handicap frequency counts (with means and standard deviations) were computed by classroom and across the total sample for:

1. the total number of happy choices in the Picture Game,
2. the total number of peer nominations on the Who Could This Be Game,
3. the number of negative nominations on the Who Could This Be Game,
4. the percent of negative nominations on the Who Could This Be Game, and
5. the total teacher rating score on the Pupil Behavior Rating Scale.

In compliance with the procedures for screening established by Bower and Lambert (1961a), these statistics were used in establishing the following screening criteria for each instrument:

1. Self rating (Picture Game) - criterion was reached if an individual's score fell one standard deviation above or below the population mean ($\leq$25 or $\geq$45).

2. Peer rating (Who Could This Be Game) - criterion was reached if the percent of negative scores was among the five highest in the class.

3. Teacher rating (Pupil Behavior Rating Scale) - criterion was reached if the total score was among the five highest in the class.

A child was "screened" if at least two out of these three scores were in the criterion region.

Analytic strategy for validity study for gifted and educable mentally retarded. Frequency counts as well as means and standard deviations were computed for the IQ scores of children in the 25 project classrooms. These scores were used to identify possible gifted and EMR children for criterion groups as well as to define a sample to be studied further. When the potential EMR and gifted criterion groups were defined, a discriminant analysis was performed to determine which screening variables, if any, differentiated between the groups. Essentially, discriminant analysis is a procedure which evaluates the ability of a set of independent variables to statistically distinguish or separate two or more sets of individuals with known group membership. In this study, discriminant analysis was used to judge the degree to which six scores obtained from the screening measures would distinguish 1) probable gifted vs. normal IQ pupils and 2) normal IQ vs. probable EMR pupils. The six scores from the screening materials used in the discriminant analysis were:

1. Teacher Rating I (Classroom Adaptation) - Subscore for four attributes from the Pupil Behavior Rating Scale which indicate the readiness of a pupil to accomplish learning tasks.

2. Teacher Rating II (Interpersonal Adjustment) - Subscore for three attributes from the Pupil Behavior Rating Scale indicating the degree to which a pupil experiences difficulties in getting along with others, is pugnacious or aggressive.

3. Teacher Rating III (Intrapersonal Adjustment) - Subscore for four attributes from the Pupil Behavior Rating Scale reflecting the degree to which a pupil experiences difficulties because he/she is immature, unhappy, friendless or withdrawn.

4. Peer Rating - Percent negative nominations from classmates on the Who Could This Be Game.

5. Self Rating I - Total number of negative nominations by self for roles on the Who Could This Be Game.

6. Self Rating II - Total number of happy pictures chosen by pupils on the Picture Game.

Results of the Validity Substudy

Forming IQ criterion groups. The major goal of the project was to explore the possibility of using the "Nonintellectual Assessment" procedures for initial screening of educable mentally retarded (EMR) and mentally gifted minors (MGM). To establish criterion groups, Slosson IQ tests were administered to identify pupils who scored in the extremes of the IQ distribution. Based on norms in use for the Stanford-Binet Intelligence Test (S-B) and procedures established by the California State Department of Education, it was decided *a priori* that those who scored 133 and above would be considered gifted and those who scored 68 and below would be considered EMR. Table 4 presents the descriptive statistics for the administration of the Slosson Intelligence Test to the project sample.

As these statistics make apparent, the mean IQ score for the project sample was higher than would be expected from a representative sample of

Table 4

Descriptive Statistics for Slosson Intelligence Test

Mean	113.09
Median	113.72
Mode	116.00
Standard Deviation	16.33
Lowest Score	72
Highest Score	175

school children like those constituting the norm groups of the S-B and WISC-R. In such samples S-B and WISC-R scores fluctuate about a mean score of 100. In the project sample the mean score was 113 and scores ranged from 72 to 175. Consequently, using the criteria* of the S-B and WISC-R, none of the participating project children would be selected as possible EMR, while 63 children, or 11 percent, would be identified as potentially mentally gifted minors (MGM). It is doubtful that such an outcome would be found if S-B or WISC-R tests had been administered instead of the Slosson IQ test.

*In California the criteria for MGM is 133 on the S-B, or 131 on the WISC-R. For EMR it is 68 on the S-B, or 70 on the WISC-R.

The problem for this study is of another sort. If the a priori criterion is used, no low IQ children are available for determing the usefulness of the screening materials in distinguishing low IQ pupils from their normal IQ peers. This dilemma made it necessary to ask whether it is more likely that the project sample was nonrepresentative or more likely that the Slosson yielded somewhat inflated scores. Two pieces of information bear on this question. The manual for the Slosson provides results of concurrent administrations of the S-B and the Slosson. For the age levels tested in the project (approximately ages 5 to 7), comparability between the S-B and Slosson is quite high. But it was the impression of the examiners for this project who were familiar with both the S-B and the Slosson that typical Slosson scores were very likely 10 to 15 points higher than would be obtained with the S-B. Although no decisive resolution of this dilemma was possible, the following course of action was taken.

For the purposes of validating the screening materials, an alternative strategy for selecting high and low IQ groups was used. In a procedure analogous to that used with the S-B and WISC-R, the IQ distribution of the project sample was used to establish cut-off scores two standard deviations above and below the obtained mean. Using this criterion, those with scores higher than 144 (N = 14) became the potentially gifted group and those with scores 81 and below (N = 14) became the low IQ group (potentially EMR). These groups were used in the discriminant analysis.

Table 5 shows the pupils identified for the high and low IQ criterion groups distributed by the demographic variables. As is apparent from this classification, proportionately fewer females than males were present in both the low and high groups. In comparison to their presence in the total sample, pupils from ethnic groups other than white were overrepresented in the low IQ group and underrepresented in the high IQ group. A similar pattern was obtained for those pupils who came from bilingual backgrounds, particularly for bilingual pupils in the "other" ethnic category. Representation across the socioeconomic levels was relatively proportional to that of the total sample for the low IQ group. For the high IQ group, pupils from the upper-socioeconomic level were overrepresented and those from the lower were underrepresented.

Discriminant analysis results. The results of the discriminant analyses are displayed in Tables 6 and 7. The statistic summarizing the differentiating power of the screening variables as a set was significant ($p < .01$) for the comparison of the low IQ group vs. those of higher IQ in the sample, but not for the high IQ group vs. those of lower IQ in the sample. Even for the low IQ dichotomy, however, the variance accounted for by the nonintellectual assessment measures was rather small (4 percent). The results did, nonetheless, favor some of the measures more than others in the screening of potentially gifted and EMR pupils.

Although the results of the analyses were not conclusive, they did permit an evaluation of the importance of each of the screening measures in the discrimination. For the high IQ vs. other dichotomy, the Teacher Rating I, Teacher Rating II, and Peer Rating measures showed the most discriminating power. High IQ pupils were viewed by their teachers as

Table 5

Distribution of Pupils Obtaining Low and High IQ Scores Summarized by Sex, Socioeconomic Level, Ethnicity and Bilingual Status

		Low IQ											High IQ										
Sex		Male					Female					Total (Percent)	Male					Female					Total (Percent)
Socioeconomic level		U	UM	M	LM	L	U	UM	M	LM	L		U	UM	M	LM	L	U	UM	M	LM	L	
Ethnicity	Bilingual Status																						
White	B	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	NB	0	1	1	0	1	0	0	0	0	0	3 *21.4*	3	1	3	4	0	1	1	0	0	0	13 *92.9*
Black	B	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	NB	0	1	0	0	0	0	0	0	0	0	1 *7.1*	0	0	0	0	0	0	0	0	0	0	0 *0.0*
Chinese	B	0	0	0	0	0	0	0	0	0	0	0 *0.0*	0	0	0	0	0	0	0	0	0	0	0 *0.0*
	NB	0	0	0	0	0	0	0	0	0	0	0 *0.0*	0	0	0	0	0	0	0	0	0	0	0 *0.0*
Filipino	B	0	0	1	0	0	0	0	0	0	0	1 *7.1*	0	0	0	0	0	0	0	0	0	0	0 *0.0*
	NB	0	0	0	1	0	0	0	0	0	0	1 *7.1*	0	0	0	0	0	1	0	0	0	0	1 *7.1*
Japanese	B	0	0	0	0	0	0	0	1	0	0	1 *7.1*	0	0	0	0	0	0	0	0	0	0	0 *0.0*
	NB	0	0	0	0	0	0	0	0	0	0	0 *0.0*	0	0	0	0	0	0	0	0	0	0	0 *0.0*
Mexican	B	0	0	0	0	1	0	0	0	0	0	1 *7.1*	0	0	0	0	0	0	0	0	0	0	0 *0.0*
	NB	0	0	1	0	0	0	0	0	1	0	2 *14.3*	0	0	0	0	0	0	0	0	0	0	0 *0.0*
Other	B	0	1	1	0	0	1	1	0	0	0	4 *28.6*	0	0	0	0	0	0	0	0	0	0	0 *0.0*
	NB	0	0	0	0	0	0	0	0	0	0	0 *0.0*	0	0	0	0	0	0	0	0	0	0	0 *0.0*
Total (Percent)		0 *0.0*	3 *21.4*	4 *28.6*	1 *7.1*	2 *14.3*	1 *7.1*	1 *7.1*	1 *7.1*	1 *7.1*	0 *0.0*	14	3 *21.4*	1 *7.1*	3 *21.4*	4 *28.6*	0 *0.0*	2 *21.4*	1 *7.1*	0 *0.0*	0 *0.0*	0 *0.0*	14

Table 6

Discriminant Function Analysis of "Nonintellectual Assessment" Measures for Low IQ Group vs. Others

Nonintellectual Assessment Measures	Low IQ (N=10) $\bar{X}$	s.d.	Others (N=478) $\bar{X}$	s.d.	Discriminant Function	Cumulative Explained Variance	Change in Variance
Teacher Rating I (adaptation)	7.60	3.29	4.73	2.68	-1.00	.02**	
Teacher Rating II (Interpersonal)	4.03	1.88	3.83	1.91	.68	.03**	.01
Peer Rating (% Negative Nominations)	60.60	23.74	45.51	20.39	-.48	.04**	.00
Self Rating II (Number happy choices)	32.30	14.57	35.07	9.75	.19	.04**	.00
Self Rating I (Number negative choices)	2.40	3.17	2.47	2.17	.05	.04**	.00

Explained variance = .04
Wilk's Lambda = .96
Canonical Correlation = .19
$X^2 = 18.58$, df = 5, $p < .01$

** $p<.01$

Table 7

Discriminant Function Analysis of "Nonintellectual Assessment" Measures for High IQ Group vs. Others

Nonintellectual Assessment Measures	High IQ (N=14) $\bar{X}$	s.d.	Others (N=478) $\bar{X}$	s.d.	Discriminant Function	Cumulative Explained Variance	Change in Variance
Teacher Rating I (Adaptation)	3.06	1.97	4.85	2.73	.97	.01*	
Teacher Rating II (Interpersonal)	3.54	1.95	3.84	1.91	-.56	.01*	.00
Peer Rating (% Negative Nominations)	37.36	15.39	46.13	20.61	.42	.02*	.01
Self Rating I (Negative choices)	2.14	1.75	2.48	2.19	.15	.02	.00
Teacher Rating III (Intrapersonal)	3.12	1.98	4.14	2.17	.06	.02	.00

Explained variance = .02
Wilk's Lambda = .98
Canonical correlation = .13
$X^2 = 8.20$, df = 5, $p < .05$

* sig <.05

having fewer difficulties with adapting in the classroom and in interpersonal areas and also received a lower percentage of negative nominations from their peers. For the low IQ vs. other dichotomy, the same three measures were the most predictive but in the opposite direction. Low IQ children were viewed as having considerably more problems in adapting to the learning demands of the classroom and in relating to others. Confirming the latter finding, the low IQ group recieved a much higher proportion of negative nominations from their classmates than did those of higher IQ.

These results are largely consistent with the results of a similar study which was recently completed (Lambert, Hartsough, and Urbanski, 1977). In that sample also, the Teacher Rating and Peer Rating measures were the two which had the most discriminating power for identifying the low IQ children. Although the Teacher Rating was one of the two best predictors for very high IQ children, as it is in the present study, higher than average Self Rating rather than Peer Rating was the other best predictor.

One outcome of the discriminant studies worth noting was the information from what is called the "classification phase" of the discriminant analysis. In this phase, all cases from the original samples are classified into the groups of interest based not on their known group membership but on the predictions derived from the nonintellectual assessment procedures as derived from the analysis phase of the discrimination problem. Classification on the basis of this prediction function is then compared to known group membership. The accuracy of this classification is often taken as an indication of the relative usefulness (predictive validity) of the variables entering the analysis.

In the two studies of interest here, 70 percent of the known cases in the low vs. all others dichotomy were correctly classified by the nonintellectual assessment materials alone. For the high IQ vs. all others dichotomy, 79 percent of the known cases were correctly classified. Using only the information from the screening variables, seven of the ten low IQ pupils and 11 of the 14 high IQ pupils would have been correctly distinguished from their peers. In both cases, however, there also would have been a substantial number of misclssifications of other pupils as possible EMR or gifted. For the gifted, 187 pupils, or 39 percent, of the total would have been misclassified, while 113, or 23 percent, of the remaining sample would have been classified as probable EMR. Obviously, additional refinements and closer analysis of the screening results are needed to improve the accuracy of the discrimination.

Results of the Screening Substudy

Of the 536 children who participated, 75, or 14 percent, were screened by the criteria given on page 20.* This result closely approximates results of other studies which have found that about 13 percent of the school population was identified as learning handicapped (EH) (Lambert, 1963). Tables 8 through 11 give a detailed characterization by sex, socio-economic level, ethnicity, and bilingual status of those who were screened by each combination of measures in the multiple cut-off procedures. As these tables indicate, each one of the screening instruments made a

*The proportion of children who are screened is a function of the strictness in application of the multiple cut-off criterion established by Bower and Lambert. A lesser number is screened from each classroom if, for example, only three rather than five of the most extreme scores on each measure is included in the cut-off region.

Table 8

Distribution of Pupils Screened by Teacher-Peer-Self Multiple Cut-off Criterion Categorized by Sex, Socioeconomic level, Ethnicity and Bilingual Status

Sex		Male						Female						Total
Socioeconomic Level		U	UM	M	LM	L		U	UM	M	LM	L		
Ethnicity	Bilingual Status													
White	B	–	–	–	–	–		–	–	–	–	–		–
	NB	0	1	2	1	0		0	0	0	2	0		6
Black	B	–	–	–	–	–		–	–	–	–	–		–
	NB	0	0	0	0	0		0	0	0	0	0		0
Chinese	B	0	0	0	0	0		0	0	0	0	0		0
	NB	0	0	0	0	0		0	0	0	0	0		0
Filipino	B	0	0	0	0	0		1	0	0	0	0		1
	NB	0	0	0	0	0		0	0	0	0	0		0
Japanese	B	0	0	0	0	0		0	0	0	0	0		0
	NB	0	0	0	0	0		0	0	0	0	0		0
Mexican	B	0	0	0	0	1		0	0	0	0	0		1
	NB	0	0	0	0	0		0	0	0	0	0		0
Other	B	0	0	0	0	0		0	0	0	0	0		0
	NB	0	0	0	0	0		0	0	0	0	0		0
Total		0	1	2	1	1		1	0	0	2	0		8

Table 9

Distribution of Pupils Screened by Teacher-Peer Multiple Cut-off Criterion Categorized by Sex, Socioeconomic Level, Ethnicity and Bilingual Status

Sex		Male						Female						Total
Socioeconomic Level		U	UM	M	LM	M		U	UM	M	LM	L		
Ethnicity	Bilingual Status													
White	B	-	-	-	-	-		-	-	-	-	-		-
	NB	1	1	5	13	4		1	1	2	2	0		30
Black	B	-	-	-	-	-		-	-	-	-	-		-
	NB	0	0	0	0	0		0	0	0	0	0		0
Chinese	B	0	0	0	0	0		0	0	0	0	0		0
	NB	0	0	0	0	0		0	0	0	0	0		0
Filipino	B	0	0	0	0	0		0	0	0	0	0		0
	NB	0	1	0	2	0		0	0	0	0	0		3
Japanese	B	0	0	0	0	0		0	0	0	0	0		0
	NB	0	0	0	0	0		0	0	0	0	0		0
Mexican	B	0	0	0	0	0		0	0	0	0	0		0
	NB	0	0	1	1	1		0	0	0	0	0		3
Other	B	0	0	0	0	0		0	0	0	0	0		0
	NB	0	0	0	0	0		0	0	0	0	0		0
Total		1	2	6	16	5		1	1	2	2	0		36

Table 10

Distribution of Pupils Screened by Teacher-Self Multiple Cut-off Criterion Categorized by Sex, Socioeconomic Level, Ethnicity and Bilingual Status

Sex		Male						Female						Total
Socioeconomic Level		U	UM	M	LM	L		U	UM	M	LM	L		
Ethnicity	Bilingual Status													
White	B	-	-	-	-	-		-	-	-	-	-		-
	NB	1	1	1	4	4		1	0	0	3	0		15
Black	B	-	-	-	-	-		-	-	-	-	-		-
	NB	0	0	0	1	0		0	0	0	1	0		2
Chinese	B	0	0	0	0	0		0	0	0	0	0		0
	NB	0	1	0	0	0		0	0	0	0	0		1
Filipino	B	0	0	0	1	0		0	0	0	0	0		1
	NB	0	0	0	1	0		0	1	0	0	0		2
Japanese	B	0	0	0	0	0		0	0	0	0	0		0
	NB	0	0	0	0	0		0	0	0	0	0		0
Mexican	B	0	1	0	0	0		0	0	0	0	0		1
	NB	0	0	0	0	0		0	0	0	0	0		0
Other	B	0	0	1	0	0		0	0	0	0	0		1
	NB	0	0	0	0	1		0	0	0	0	0		1
Total		1	3	2	7	5		1	1	0	4	0		24

Table 11

Distribution of Pupils Screened by Peer-Self Multiple Cut-off Criterion Categorized by Sex, Socioeconomic Level, Ethnicity and Bilingual Status

Sex		Male					Female					Total
Socioeconomic Level		U	UM	M	LM	L	U	UM	M	LM	L	
Ethnicity	Bilingual Status											
White	B	-	-	-	-	-	-	-	-	-	-	-
	NB	0	1	3	0	0	0	0	1	1	0	6
Black	B	-	-	-	-	-	-	-	-	-	-	-
	NB	0	0	0	0	0	0	0	0	0	0	0
Chinese	B	1	0	0	0	0	0	0	0	0	0	1
	NB	0	0	0	0	0	0	0	0	0	0	0
Filipino	B	0	0	0	0	0	0	0	0	0	0	0
	NB	0	0	0	0	0	0	0	0	0	0	0
Japanese	B	0	0	0	0	0	0	0	0	0	0	0
	NB	0	0	0	0	0	0	0	0	0	0	0
Mexican	B	0	0	0	1	0	0	0	0	0	0	1
	NB	0	0	0	0	0	0	0	0	0	0	0
Other	B	0	0	0	0	0	0	0	0	0	0	0
	NB	0	0	0	0	0	0	0	0	0	0	0
Total		1	1	3	1	0	0	0	1	1	0	8

contribution in the final identification process. The score combinations which selected pupils on the basis of extreme scores on the teacher and peer rating instruments and the teacher and self rating instruments identified the majority of those screened (approximately 80 percent). Of the remaining 20 percent, half were screened with extreme scores on the peer and self rating instruments and the other half with extreme scores on all three measures. Table 12 presents a compilation by the major demographic indicators of the numbers and percent screened regardless of score combination. In the total sample, boys and girls were almost equally represented. However, among those screened, 80 percent were boys and 20 percent girls, a discrepancy confirming once again the greater susceptibility of boys to problems in elementary school.

The preponderance of those screened as EH were from families of middle and lower middle class backgrounds. As compared to their proportions in the total sample, these two levels were very slightly overrepresented in the screened sample. Pupils of upper middle class background were slightly underrepresented among those screened, while those at the extremes of the social class continuum were present in the screened group in the same proportions as in the total sample.

With regard to the other demographic variables, pupils from all the major ethnic groups, excluding Japanese, were present among those screened and in proportions nearly identical to their representation in the total sample. Similarly, the proportions of screened pupils who came from families where the dominant language was not English were small and again consistent with their representation in the total group.

Table 12

Frequencies of Pupils Screened by Any of Multiple Cut-off Criteria Summarized by Major Demographic Variables

Sex		Male					Female					Total (Percent)	
Socioeconomic Level		U	UM	M	LM	L	U	UM	M	LM	L		
Ethnicity	Bilingual Status												
White	B	–	–	–	–	–	–	–	–	–	–	–	
	NB	2	4	11	19	8	2	1	3	7	0	57	*75.0*
Black	B	–	–	–	–	–	–	–	–	–	–	–	
	NB	0	0	0	1	0	0	0	0	1	0	2	*2.6*
Chinese	B	1	0	0	0	0	0	0	0	0	0	1	*1.3*
	NB	0	1	0	0	0	0	0	0	0	0	1	*1.3*
Filipino	B	0	0	0	1	0	1	0	0	0	0	2	*2.6*
	NB	0	1	0	3	0	0	1	0	0	0	5	*6.6*
Japanese	B	0	0	0	0	0	0	0	0	0	0	0	*0.0*
	NB	0	0	0	0	0	0	0	0	0	0	0	*0.0*
Mexican	B	0	1	0	1	1	0	0	0	0	0	3	*4.0*
	NB	0	0	1	1	1	0	0	0	0	0	3	*4.0*
Other	B	0	0	1	0	0	0	0	0	0	0	1	*1.3*
	NB	0	0	0	0	1	0	0	0	0	0	1	*1.3*
Total (Percent)		3 *4.0*	7 *9.0*	13 *17.1*	26 *34.2*	11 *14.4*	3 *4.0*	2 *3.0*	3 *4.0*	8 *10.5*	0 *0.0*	76	

Subtotals: Male 60 (80) Female 16 (20)
SES: Upper 6 (08) Upper Middle 9 (12) Middle 16 (21)
Lower Middle 34 (45) Lower 11 (14)

Incidental Findings

Congruence of teacher perception and project results. Twenty-three of the 25 participating teachers could be reached for the feedback phase of the project. Each feedback session lasted approximately 45 minutes to one and one-half hours, depending upon the number of pupils to be discussed and the kinds of questions and issues raised by the teacher. The feedback form provided space for the person giving the feedback to write comments from the teacher about her agreement or disagreement with the findings. No attempt was made by project personnel to encourage responses from teachers in the direction of agreement with project findings. Results were reported in a framework of planning future programs or interventions for the pupils, and it was the impression of those doing the reporting that teacher reactions were candid.

In all, reports were made on some 300 children. Thirty-six of these reports pertained exclusively to the results of administration of the ability test, that is, the discussion of children who were potential candidates for gifted programs. In about 50 percent of the cases where high scores were reported, teachers expressed surprise at the results. Often they indicated that the pupil's in-class performance gave them no reason to suspect such good learning ability. Teachers concurred more frequently about the reports for low scoring students, although far from unanimously. They perceived some 20 to 30 percent of those with low scores as experiencing relatively few problems with the academic requirements of kindergarten or first grade.

The remaining 264 reports were about the outcomes obtained with the screening materials. For these reports, agreement between project results

and teacher perception was quite good. In only 21 instances (8 percent of the 264 reports) was there lack of congruence and in many of these cases the discordant results could be attributed to the pupil being new to the class and therefore relatively unknown to both the teacher and peers at the time of administration of the materials.

Comparison of results of screening with teacher nominations made prior to administration of materials. In an informal attempt to compare the efficiency of the assessment process with existing information which might be used for identification purposes, we asked teachers in three of the participating districts (Ceres, Empire City, Modesto) to nominate children as gifted, EMR and EH candidates before the study began. Comparison of nominations with the measures collected would enable us to infer the potential of the procedures in identifying pupils who might be overlooked by teachers in the traditional referral process. Seven teachers submitted 28 nominations. Of these nominations, 11 pupils were nominated for gifted, two as possible EMR, and 15 as EH candidates.

Among those nominated for the gifted and EMR classifications, results from the administration of the Slosson Intelligence Test yielded scores which showed no overlap with initial teacher nominations. For the EH category, congruence between teacher nominations and the nonintellectual assessment procedure was greater. Six of the 11 pupils nominated by teachers were screened by the multiple cut-off procedures of the nonintellectual assessment process as needing further evaluation. In total, 25 pupils were screened by the nonintellectual assessment procedures in the seven classrooms as having possible learning and/or behavior problems.

Relying on teacher referral alone would mean that only six of these 25 vulnerable pupils would be viewed as eligible for additional services.

Comparison of results of administration of ability test in native language and in English. Twenty-two pupils in the project were administered the Slosson Intelligence Test in languages other than English. Of these, 13 were tested in Tagalog and nine in Spanish. Since examiners and funds were available, it was decided to make an informal assessment of the comparability of scores obtained on the SIT in the native language and in English for the Tagalog-speaking group. The scores for both test administrations and the resulting point differences are displayed in Table 13. With only one exception, all scores for the English administration of the SIT fell below the mean score (113.09) obtained by the project sample. Five of the 13 Tagalog-speaking children had scores one and one-half standard deviations below the sample mean (89 and below) for the English administration. Results for the Tagalog administration show an improved picture. Three of the children scored above the sample mean and only two scored in the low range. A review of the score differences for the two administrations revealed that scores increased in varying degrees for nine of the 13 children when the test was administered in Tagalog. In five cases the increase was substantial, the scores having improved by more than one standard deviation. With the Tagalog administration, one child achieved a score (141) which would signal recommendation for further evaluation for a gifted program. For three children scores decreased, though not substantially, when the test was administered in Tagalog. These may be children who have facility in

Table 13

Scores Obtained for Pupils Administered Ability Test in English and in Tagalog

Pupil	Score for Tagalog Administration	Score for English Administration	Difference Between Administrations
1	114	90	24
2	110	93	17
3	129	112	17
4	141	127	14
5	101	87	14
6	100	89	11
7	92	86	6
8	103	99	4
9	100	99	1
10	101	101	0
11	93	96	-3
12	78	82	-4
13	82	87	-5

neither language. In general, these findings strongly suggest that ability test results for children whose dominant language is not English are questionable when the score is from a test administered in English.

Summary and Recommendations

The primary purpose of the study was an evaluation of the "Nonintellectual Assessment" procedures for the preliminary identification of gifted and EMR pupils early in their educational experience. The early identification of children in need of special learning assistance might help prevent several years of failure or frustration for such pupils by making school personnel aware of their instructional needs. The question posed by the study was whether or not the nonintellectual assessment procedures were a more efficient or accurate approach to early identification than the historical teacher-referral process. Pupils identified by the procedures could become targets for intensive instructional intervention as well as candidates for more extensive review within the district or county.

A second facet of the study was the utilization of the screening materials to identify those pupils whose emotional or behavioral symptoms predispose them to less extreme forms of learning handicaps. The identification of "educationally handicapped", or learning handicapped, pupils was the purpose for which the screening materials were originally developed and the function they have traditionally served. We wished to obtain descriptive statistics about the rate of occurrence and the nature of the group of pupils identified in this way, as well as to provide feedback to the teachers who participated in the study. Information about children who were vulnerable to potential behavioral and learning difficulties was shared with each teacher.

A third aspect was to assess the degree of congruence between our results and teacher perceptions of their pupils at the end of the school

year. Prior to the inception of the project, a small sample of the teachers nominated pupils who they felt were candidates for gifted, educable mentally retarded, and learning handicapped assessment. A comparison of the screening process with teacher nominations for referral was another component of the study.

And last, we explored the comparability of pupil scores on some of the project instruments when the assessment was in both English and a pupil's native language.

Are the screening materials a more efficient approach to early identification of educable mentally retarded and gifted pupils than the traditional teacher referral process? There is, unfortunately, no simple answer. In general, the screening materials seem to "over-predict" or include too many youngsters, while teacher nominations or reviews seem to "under-predict" or exclude many children who might benefit from educational interventions.

The screening instruments alone identified 79 percent of the children in the 25 project classrooms whose scores on the IQ test suggested them as candidates for a gifted program. However, the screening materials also identified as gifted a substantial proportion of the total sample (39 percent) whose IQ scores did not suggest them as candidates for a gifted program, but who were rated by their teachers as having characteristics associated with gifted children: few difficulties adapting in the classroom and in interpersonal relationships, and few negative nominations by peers. Among the teachers who did early nominations, none of the children whose IQ scores suggested gifted potential were nominated as gifted by their

teachers. When pupils with "excellent ability" were identified to teachers at the end of the study, teachers seemed surprised about 50 percent of the time.

Of the children whose scores on the IQ test indicated possible candidacy for educable mentally retarded programs, 70 percent were identified by the screening materials alone. As with the gifted, the screening materials identified a large proportion of the total sample (23 percent) as having the same characteristics that distinguished the the potential EMR candidates: rated by their teachers as having many difficulties adapting in the classroom and in interpersonal relationships, and associated with negative behaviors by their classmates. Again, none of the children whose IQ scores suggested EMR potential were nominated by their teachers early in the study. When pupils with "low ability indicating possible learning problems" were identified to the teachers at the end of the study, teachers seemed to agree about 70 or 80 percent of the time.

The historically validated use of the screening materials has been for the identification of children for possible placement in learning handicapped programs. Six of the 11 pupils nominated by seven of the project teachers as probable learning handicapped candidates were also identified by the screening materials. However, nineteen other children were also identified by this known and proven application of the materials as probable candidates. The congruence between the initial teacher nominations and children identified by the screening materials was about 55 percent. However, relying on teacher referral alone would mean that only

six of the 25 vulnerable pupils would be viewed as eligible for additional services.

Was this population similar to others in the numbers and the characteristics of the educationally handicapped children who were identified? About 14 percent of the pupils in the study were screened as learning handicapped. This closely approximates results of the earlier studies. In general, those screened as learning handicapped reflected the social class continuum in the sample. Pupils from all the major ethnic groups, excluding Japanese, were present among those screened in proportions nearly identical to their representation in the total sample. Similarly, the proportions of screened pupils who came from families where the dominant language was not English were small and consistent with their representation in the total group. As in other studies, the results of the screening materials seem less vulnerable to socioeconomic status and ethnic or bilingual background than do IQ scores.

Did the testing of non- or limited English-speaking youngsters in their native languages make a difference in the results? We made an informal assessment of the comparability of scores obtained on the IQ test administered in the native language and in English for a group of 13 Tagalog-speaking pupils. The results strongly suggest the need for ability testing in their native languages for children whose dominant language is not English. The general tendency for the IQ scores from tests given in both English and Tagalog was to move from the medium level to the high range when the test was given in Tagalog. Low IQ

scores tended to stay in place in either language. The results suggest that ability testing in a child's native language may be particularly critical for the identification of gifted children.

Of the ten minority children in the total sample who had low IQ scores, seven were bilingual. That suggests that low ability bilingual children need to be identified early and to be given intensive language skills training. It is very possible that they should not be considered for special placement until several years of such intervention has occurred.

Conclusions

This study of the validation of a procedure for the early screening of gifted and educable mentally retarded pupils should be regarded as exploratory. The results, however, are quite intriguing and suggest the need for future research with these or similar screening procedures. When compared to IQ scores, the nonintellectual assessment materials appear to be far too inclusive as screening devices for probable gifted and probable educable mentally retarded pupils. Yet teacher judgment appears to be far too restrictive, particularly in the case of probable gifted children, when compared to IQ scores.

As educators, we are faced with some difficult questions. Is it better to overidentify and attempt to provide enriched educational programs for a number of children, some of whom may not fully profit from an enriched program or meet the criteria of giftedness in later years? Or should we continue with a system in which many gifted children are not identified in their early years of schooling and find school to be

boring and frustrating, with loss of potential for later years? Is it preferable to overidentify and provide intensive instructional programs for some children who may not be EMR, as defined by later criteria? Or should we continue to place in EMR programs only those children for whom we have positive identification and leave others in regular school programs where they face frustration because they cannot understand basic concepts and lack certain skills necessary to compete in regular classrooms?

It would be interesting to determine what might happen to the academic achievement and aptitude scores if all pupils who were identified by the screening materials as potentially gifted were exposed to several years of enriched curricular experiences. It would also be interesting to attempt intensive intervention for all children who were screened as candidates for educable mentally retarded programs. The costs (money, frustration, possible stigma, self-fulfilling prophecies) and benefits (increased motivation, self-esteem, tailored curriculum, increased learning) to both the individual and society of such alternative approaches should be contrasted to our current practices.

References

Bower, E.M. & Lambert, N.M. A process for in-school screening of emotionally handicapped children: Manual for school administrators and teachers. Princeton, NJ: Educational Testing Service, 1961a.

Bower, E.M. & Lambert, N.M. A process for in-school screening of children with emotional handicaps: Technical report for school administrators and teachers. Princeton, NJ: Educational Testing Service, 1961b.

Buros, O.K. The Seventh Mental Measurements Yearbook. Highland Park, NJ: The Gryphon Press, 1972.

California State Department of Education. California State Testing Program, 1971-72 and 1972-73: Profiles of School District Performance. Sacramento: California State Department of Education, 1974.

California State Department of Education. 1977 California Public School Directory. Sacramento: California State Department of Education, 1977.

Lambert, N.M. The development and validation of a process for screening emotionally handicapped children. USOE Cooperative Research Project 1186. Sacramento: California State Department of Education, 1963.

Lambert, N.M., Hartsough, C., Converse, F.L. & Converse, E.K. APPLE: An annotated pupil information system. Educational Technology, 1971, 11 (11), 32-36.

Lambert, N.M. Intellectual and nonintellectual predictors of high school status. The Journal of Special Education, 1972, 6(3), 247-259.

Lambert, N.M. The development of instruments for the nonintellectual assessment of effective school behavior. Technical Report Supplement. University of California, Berkeley. Unpublished manuscript, 1974.

Lambert, N.M., Hartsough, C.S. & Urbanski, C. Instituting new early screening procedures for the Newark Unified School District. Final Report. University of California, Berkeley, 1977.